Research review: A computer-based diagnostic model for individual case review

A previous study found that Iliad, a diagnostic expert system, detects diagnostic errors missed by peer review organization (PRO) review. That study used volunteer physicians from an institution as gold standard reviewers, however. The article discusses a second experiment employing Utah PRO (UPRO) review physicians as gold standards. Iliad was compared with the Unified Clinical Data Set used by the UPRO and was found to detect otherwise unsuspected diagnostic errors. The confirmation rate of Iliad flags was much higher in the earlier study, however. No agreement was found between institution and UPRO physicians, but there was agreement between a unique physician (who was both an institution and UPRO physician) and each of the two groups. Because Iliad screens for potential diagnostic errors to be confirmed or denied by gold standard physician review, the different types of physicians in the two experiments might have been the cause. Key words: *diagnostic error, expert system, Iliad, peer review organization, quality review, Unified Clinical Data Set.*

Lee Min Lau, M.B.B.S., Ph.D.
Chief, Quality Assurance and Outcomes Research
Department of Hospital and Clinical Informatics
University of Utah Health Sciences Center

Homer R. Warner, Ph.D., M.D.
Chair and Professor
Department of Medical Informatics
University of Utah

A. Lloyd Poulsen, M.S.P.H., M.D.
Medical Director
Utah Peer Review Organization
Salt Lake City, Utah

THE GOAL OF QUALITY assurance is to improve health care delivery through the monitoring and analysis of patient management strategies.[1–4] The detection of a quality problem leads to feedback to the caregiver and thus improvement in future patient care. This is the concept underlying the review procedures employed by peer review organizations (PROs) for Medicare cases. Ideally, the monitoring and analysis for quality problems is performed by expert physicians through peer review. The high cost of physician review, however, combined with the great volume of patient cases screened by the PROs prohibits a physician review of all cases. Therefore, the PROs sample a subset of cases to go through a preliminary nurse screening. The nurse reviewers apply generic quality screening rules and flag cases containing potential quality

The authors thank the staff and physicians of the Utah Peer Review Organization for their help and cooperation. We also thank Dr. D. Bjorkman, Dr. B. Bray, Dr. P. Haug, Dr. M. Lincoln, Dr. A. Renzetti, and Dr. F. Tyler of the University of Utah. Supported in part by grant 1-R03-HS06947-01 from the Agency for Health Care Policy and Research.

Top Health Inform Manage, 1995, 15(3), 67–79

problems, which are then referred for an expert physician review.[5,6]

The Utah PRO (UPRO) reviews about 10,000 Medicare cases per year (approximately a quarter of all Utah Medicare claims).[7] From April 1989 to March 1990 18.0 percent of the Medicare cases reviewed were flagged by the nurse reviewers as containing potential quality problems and were referred to physician reviewers. The physician review confirmed quality problems in only 5.5 percent of these 18 percent, or 1 percent of the original 10,000 cases. This is a surprisingly low figure compared with the much higher base rate of quality problems indicated by a substantial body of medical literature, which ranges from 3 percent to 42 percent.[8-15] Other PROs have reported quality problem rates between 0.3 percent and 4.6 percent, and the national PRO average was reported to be 1.57 percent.[16,17]

The PROs' lower problem detection rate compared with the rates found by other researchers may be due to underdetection of quality problems arising from diagnostic errors. The present PRO review begins with identification of probable quality problems by nurse reviewers. Only then will these cases be referred to physician reviewers, with specific questions from the nurses requiring answers from the physicians. The physician review, directed by the concerns raised in the nurse review, may also fail to detect diagnostic errors. Nurses do not receive specific training in diagnosis other than diagnosis-related group (DRG) validation, and the generic quality screens used in the nurse review focus almost entirely on therapeutic and documentation errors. Most cases referred to physicians for DRG validation do not raise quality concerns on the part of the physicians. Previous research indicates that diagnostic errors are certainly present.[8-11] For instance, one study reported a missed or delayed diagnosis in 10 percent of a series of inpatient cases covering five DRGs in internal medicine.[10]

Increasing the problem detection rate solely by using physician peer review is unlikely to be practical given the large number of cases to be reviewed. Expert systems (computerized diagnostic systems) may provide an effective alternative means to detect quality problems. For instance, Stewart and colleagues described a computerized quality assurance system to assist a full-time quality assurance officer.[18] The system contained quality screening criteria for emergency department case review. The number of patient cases referred for investigation of questionable care rose from a preimplementation rate of 5 patient care errors per month to 35 per month.

In a previously presented experiment, we studied the potential use of a medical expert system called Iliad to detect diagnostic errors that lead to important quality problems in patient care.[19] Iliad is designed to act as a diagnostic consultant.[20] Given the same set of patient data, a difference in opinion between Iliad and the attending physician regarding the diagnosis may indicate the presence of a diagnostic error. These cases can then be referred for physician review with the appropriate questions about diagnosis. In our previous experiment, for 100 Medicare cases flagged by the UPRO nurse review as containing potential problems and that had a diagnosis recognized by Iliad, Iliad review found 17 of the cases to contain a diagnostic error, whereas the PRO review did not find any diagnostic errors at all. The experiment had some limitations, however. Because UPRO did not keep the paper charts of those cases found to be problem free by the nurse review, we restricted our case selection to those cases flagged by the nurse review as containing potential problems and for which paper charts were available in UPRO. We recognized that these cases might have had a higher incidence of diagnostic errors, some of which might have given rise to the therapeutic and management problems flagged by the UPRO nurse

review. The results of this experiment were thus not generalizable to all Medicare cases (with a diagnosis recognized by Iliad) Also, because of budgetary constraints, we used volunteer physicians from our institution, the University of Utah School of Medicine, as gold standard reviewers for the Iliad review. Thus, reviewer bias could not be excluded.

Therefore, a second experiment was undertaken, with help from the Agency for Health Care Policy and Research, employing the UPRO's own review physicians in the gold standard review of the cases flagged by Iliad. In addition, we decided to compare Iliad's screening performance with that of the Unified Clinical Data Set (UCDS), a computer program designed by the Health Care Financing Administration (HCFA), because UPRO was one of seven UCDS test sites.[21,22] The UCDS is a rule-based expert system that contains the present generic quality screens from the HCFA. The PRO nurse reviewers enter clinical data requested by the UCDS, including such items as vital signs, laboratory results, and procedures, but not many history or physical findings. The UCDS system will analyze the data and apply multiple HCFA quality screens. These screens are not specifically designed to detect diagnostic errors that may lead to quality problems, however. We thus anticipated that the UCDS nurse review would perform similarly to the manual nurse review with regard to diagnostic error detection.

METHODS AND PROCEDURES

Iliad is an expert system designed for diagnosis in internal medicine and other related fields with which a primary care physician must be familiar.[20] The Iliad version used in the experiment recognized more than 6,300 medical findings and 1,350 diagnostic conditions.

For its phase I UCDS testing, UPRO randomly selected 6 percent of all Utah Medicare hospitalizations in the period from February 1991 to July 1991, a total of 1,037 cases. In the UCDS review, the nurse screening was performed by UPRO nurses using the UCDS program, and potential problem cases flagged by the program were referred to UPRO physician reviewers. The remaining cases would be considered problem free and would not be referred for physician review. The physician review of the referred cases would confirm problems in some cases and deny the presence of problems in the rest. This was similar to the manual UPRO review process. An important difference was that all 1,037 cases, including those classified by the UCDS nurse review as problem free, were kept in UPRO and available for use in our experiment. Therefore, unlike the earlier experiment, the results of this second experiment could be generalizable to all Medicare cases (suitable for Iliad review).

Each of the 1,037 cases was examined to determine whether it was suitable for Iliad review. This was to provide an indication of the proportion of Medicare cases to which Iliad could be applied. To be eligible for the experiment, the principal diagnosis in the case had to be one contained in Iliad's knowledge base. Also, the case could not be completely void of diagnostic work-up information (history, physical examination findings, and laboratory test results). According to the above criteria, of the 1,037 cases, 666 (64.2 percent) were found to be suitable for Iliad review.

A two-step random selection process was used to arrive at the cases used in the experiment. First, 500 cases were randomly selected from among the 666 cases found to be suitable for Iliad review. These 500 cases consisted of two groups: those considered problem free by the UCDS nurse review (called non–UCDS-nurse flagged) and those considered to have potential quality prob-

lems by the UCDS nurse review and referred for physician review (called UCDS-nurse flagged). From the non–UCDS-nurse flagged cases, a random selection of 50 cases was made for Iliad review. It was felt that, because diagnostic errors led to quality problems in management, diagnostic errors would be more likely to be found in the UCDS-nurse flagged cases. If comparing the results of the Iliad review for the UCDS-nurse flagged cases and the non–UCDS-nurse flagged cases showed this expectation to be fulfilled, then it might be more cost effective to use Iliad only in those cases already flagged by the PRO nurse review. To conserve the limited resources available for this experiment, we opted to review a smaller number of randomly selected non–UCDS-nurse flagged cases but to review all UCDS-nurse flagged cases. At the end of the experiment, it was discovered that 4 of the 50 non–UCDS-nurse flagged cases were initially mistakenly classified as such and had actually been flagged for physician review by the UPRO nurses. Thus 46 randomly selected non–UCDS-nurse flagged cases remained for this second experiment, of a total of 220 non–UCDS-nurse flagged cases among the 500 cases. The remaining 280 of the 500 cases were the UCDS-nurse flagged cases, and all were used in the experiment.

Iliad's consultation mode was used to review these 326 cases. In this mode, the case findings were entered into Iliad, which then provided a list of differential diagnoses. For the experiment, a general practitioner reviewed the patient record and entered into Iliad the case findings obtained from the history, physical examination, and investigative laboratory tests before treatment. Iliad then generated a list of the top 20 differential diagnoses ranked according to the diagnostic certainty (probability) that Iliad assigned to each diagnosis. The attending physician's discharge diagnoses from the discharge summary were then compared with Iliad's diagnoses for the case.

For the experiment, a general practitioner reviewed the patient record and entered into Iliad the case findings obtained from the history, physical examination, and investigative laboratory tests before treatment.

A potential diagnostic error was identified when there was a discrepancy between the attending physician's list of discharge diagnoses and the corresponding list provided by Iliad for the case. A discrepancy was defined as one of the following conditions:

- a diagnosis on the physician's list did not exceed 20 percent probability on Iliad's list (a potential unlikely diagnosis)
- a diagnosis exceeding 80 percent probability on Iliad's list did not appear on the physician's list (a potential missed diagnosis)

An identified discrepancy between the attending physician's diagnostic list and Iliad's would flag the case as requiring a gold standard physician review. For this experiment, the gold standard review was performed by UPRO review physicians. Letters inviting participation were sent out to a random selection of UPRO review physicians, and six agreed to participate. All six UPRO physicians were either general internists or family practitioners. None of them was a University of Utah physician. The principal investigator was not aware of their identity, and the UPRO physicians were also not told about Iliad, in an effort to avoid the positive reviewer bias that might have been present in the earlier experiment. The gold standard review form sent together with the patient record to these reviewers

contained one or more questions in this format: "The diagnosis of XXX was made in the case. Is this a diagnostic error?" or "The diagnosis of YYY was not made in the case. Is this a diagnosis missed by the attending physician?" The standard PRO weighing criteria for severity scoring of each quality problem were also used by the review physicians as follows:[7]

- care appropriate, without quality problems
- level I problem (score 1), without potential for significant adverse effects on the patient
- level II problem (score 5), with potential for significant adverse effects on the patient
- level III problem (score 25), with significant adverse effects on the patient

Each case flagged by an Iliad nurse review was referred to one of these six physicians through UPRO. The investigator was blinded to the case assignments, which were handled totally by UPRO. The review form was also returned from UPRO with the physician identifier removed.

RESULTS AND DISCUSSION

Patient demographics

The experiment used 326 Medicare cases. The ages of the 326 patients ranged from 34 to 103 years (average, 75 years). Of the 326 patients, 165 were women (50.6 percent). The length of stay ranged from 1 to 16 days (average, 5.2 days). The admissions were to 36 of a total of 54 hospitals in Utah. The ten most common principal diagnoses are listed in Table 1.

Results of UCDS and Iliad reviews

The test population comprised 500 cases randomly selected in two steps from a random sampling of Utah Medicare inpatient cases in a 6-

Table 1. The ten most common diagnoses in the cases used

Rank	Diagnosis	Number of cases
1	Congestive heart failure	34
2	Pneumonia	30
3	Cerebrovascular accident/transient ischemic attack	22
4	Angina	17
5	Acute myocardial infarction	16
6	Cardiac dysrhythmia	15
7	Malignancy	13
8	Peptic ulcer disease / gastrointestinal hemorrhage	10
9	Chest pain	10
10	Cholecystitis	9

month period containing a diagnosis recognizable by Iliad. These 500 cases were all sent through the UCDS nurse review, and 280 were flagged for UCDS physician review. The results of UCDS review are summarized in Table 2. For this test population, four of the quality problems found by UCDS review were diagnostic in nature, hence giving a diagnostic error rate (DER) of 4 of 500 (0.8 percent).

All 280 UCDS-nurse flagged cases and 46 of the remaining 220 non–UCDS-nurse flagged cases were sent through the Iliad review. Based on the previously described 80%/20% threshold criterion for discrepancy between Iliad's and attending physicians' diagnoses, 85 cases in the UCDS-nurse flagged group and 3 cases in the non–UCDS-nurse flagged group were flagged. Each of these flagged cases was referred for a gold standard physician review by one of six UPRO physicians recruited for the experiment. The UPRO physicians were asked to review cases for diagnostic errors and to rate the quality problem according to the PRO severity classification. The gold standard physician review results are also summarized in Table 2.

Generalization of Iliad review results

The test population comprised the 500 cases randomly selected from a random sampling of

Table 2. Results of UCDS and Iliad reviews

Review process	UCDS	Iliad
Number of cases sent through nurse review	500	326
Number of cases flagged by nurse review	280	88
Number of cases with appropriate care	269	65
Number of cases with quality problems	11	23
Level I (score 1)	1	20
Level II (score 5)	10	3
Level III (score 25)	0	0
Number of diagnostic errors	4	23
Average severity score of cases with problems	4.6	1.5
Average severity score of diagnostic errors	5.0	1.5
DER of cases reviewed	0.8%	7.1%
DER projected for test population	0.8%	6.0%
TPR of cases reviewed	3.9%	26.1%
TPR of nurse review projected for test population	3.9%	30.6%

Utah Medicare inpatient cases during a 12-month period and containing diagnoses recognizable by Iliad. Therefore, the results of this experiment were generalizable to all Medicare inpatient cases with a diagnosis recognized by Iliad. The projected DER for the test population is shown in Tables 2 and 3. The difference between the DER found by Iliad review of the UCDS-nurse flagged group and the DER in the non–UCDS-nurse flagged group was not statistically significant by the *z*-test for difference between two independent proportions.[23] Thus if Iliad review is to be used to detect diagnostic errors, it should be applied to all cases with a diagnosis recognized by Iliad.

DERs and true positive rates

Iliad review resulted in a significantly higher DER as projected for the test population compared with the UCDS review ($p < .001$, McNemar's X^2 test for comparing proportions in two paired groups, $\alpha = .05$ for one-tailed test). It was not surprising that the UCDS review was unable to detect the diagnostic errors found by Iliad. The UCDS review, as the computerized form of PRO review, focused on problems of treatment, documentation, and discharge. On the other hand, Iliad review focused solely on diagnostic errors. The Iliad nurse review also has a significantly higher true positive rate (TPR) than the UCDS nurse review ($p < .001$, z test for difference between two independent proportions, $\alpha = .05$ for one-tailed test), reflecting less waste in having to pay for unnecessary, expensive physician review (see Table 2).

Quality problems found by the reviews

The quality problems found by UCDS review are summarized in the box titled "Quality Prob-

Table 3. Generalization of Iliad review results to test population ($N = 500$)

UCDS-nurse flagged cases ($n = 280$)	Non–UCDS-nurse flagged cases ($n = 220$)
280 cases sampled for Iliad nurse review	46 cases sampled for Iliad nurse review
83 cases with Iliad nurse review flags	3 cases with Iliad nurse review flags
21 cases with problems (18 level I, 3 level II)	2 cases with problems (both level I)
TPR for this group: 21/85 = 24.7%	TPR for this group: 2/3 = 66.7%
DER for this group: 21/280 = 7.5%	DER for this group: 2/46 = 4.3%
	Projected cases with problems in group: 4.3% x 220 = 9 cases
	Projected cases with flags for whole group: 9/66.7% = 13 cases
Number of problems projected for entire sample: 21 + 9 = 30 cases	
Projected DER for test population: 30/500 = 6%	
Projected TPR for test population: 30/(85 + 13) = 30.6%	

Quality Problems Found in the UCDS Review

Level I problem: 1 case

- Failure to perform a test

Level II problem: 10 cases

- Failure to diagnose pulmonary embolus: 1 case
- Failure to diagnose gouty arthritis: 1 case
- Failure to diagnose azotemia: 1 case
- Failure to diagnose pulmonary infection/fluid overload: 1 case
- Failure to perform a test: 3 cases
- Medical stability not ensured at discharge: 1 case
- Details not known: 2 cases

lems Found in the UCDS Review." Two of the UCDS physician review worksheets were not available, so that details of two level II problems were not known. Interestingly, four diagnostic errors were found among the 11 quality problems detected by UCDS review. The questions raised in the UCDS nurse review and the physician reviewers' responses for those four cases are summarized in Table 4. The 23 quality problems found by the Iliad review are presented in the box titled "Quality Problems Found in the Iliad Review." The results of the UCDS review and the Iliad review were compared to determine whether the same quality problems were detected by the two independent processes. None of the cases was found to have problems by both reviews. Thus Iliad review did not detect the four diagnostic problems found by UCDS review.

Comparison of two experiments

The Iliad review results (DER and TPR) differed considerably between this experiment and the earlier study (Table 5). Although the two experiments also differed in test population and sample size, it was thought that the most significant factor contributing to the difference in Iliad performance was the gold standard physician review. The DER found by Iliad review, as well as the TPR of the Iliad nurse review, depended

Table 4. Diagnostic errors found in the UCDS review

Nurse review concern	Physician review judgment
Medical stability not ensured at discharge	No problem with stability at discharge, but diagnosis of pulmonary embolus missed
Antibiotics started without culture	Agreed, but diagnosis of gouty arthritis missed
Chest radiograph indicated but not done	Agreed, but diagnosis of pulmonary fluid overload missed
Premature discharge with raised blood urea nitrogen and creatinine levels	Agreed, but diagnosis of azotemia missed

Quality Problems Found in the Iliad Review

Level I problem: 20 cases

- Failure to diagnose hypertensive heart disease: 1 case
- Failure to diagnose sporadic hypertriglyceridemia: 2 cases
- Failure to diagnose multiinfarct dementia: 1 case
- Failure to diagnose prerenal azotemia: 1 case
- Failure to diagnose urinary tract infection: 1 case
- Unlikely diagnosis of asthma: 1 case
- Unlikely diagnosis of congestive heart failure: 5 cases
- Unlikely diagnosis of diabetes mellitus: 1 case
- Unlikely diagnosis of renal failure: 2 cases
- Unlikely diagnosis of bilateral pneumonitis: 1 case
- Unlikely diagnosis of angina pectoris: 1 case
- Unlikely diagnosis of urinary tract infection: 2 cases
- Unlikely diagnosis of acute myocardial infarction: 1 case

Level II problem: 3 cases

- Failure to diagnose acute myocardial infarction: 1 case
- Failure to diagnose acquired hyperlipidemia: 1 case
- Unlikely diagnosis of gastritis/duodenitis: 1 case

largely on the judgment of the gold standard physician reviewers. The differences in Iliad review DER and TPR between the two experiments were both statistically significant ($p < .003$, two-tailed test) when tested by the z test for two independent proportions.[23]

There were a few reasons that could have contributed to the much better performance of the Iliad review, as judged by the gold standard review, in the first experiment. The four gold standard reviewers in the first experiment were all volunteer University of Utah physicians familiar with Iliad research. Hence they could have had a positive bias for Iliad because it was not possible to blind them to Iliad's role in the experiment. In addition, these four physicians were specialists (a gastroenterologist, a cardiologist, a pulmonary physician, and an endocrinologist) with teaching responsibilities in the University of Utah School of Medicine. None of them was on UPRO's panel of review physicians. They thus might take diagnostic errors more seriously and be less inclined to accept a faulty diagnostic process.

On the other hand, the six UPRO physicians used as gold standard reviewers in the second experiment were all general internists or family practitioners experienced in performing PRO review. Although in their gold standard review for the experiment these physicians were not aware that the questioned diagnoses were suggested by Iliad, there might have been a negative bias against this diagnostic review as yet another PRO regulatory activity to complicate the work of physicians. Also, none of these UPRO physicians was teaching at the University of Utah. They might not have thought of diagnostic error as serious as long as patient management and outcome were adequate. This is supported by the discovery that 17 of the 20 level I problems were marked as "diagnostic error present" and "care appropriate," an apparent contradiction. In the PRO severity classification, a level problem is one with no potential adverse effect on the patient (e.g., a lack of documentation). When a quality

Table 5. Comparison of the Iliad reviews in the two experiments

Experiment result	Experiment 1: Iliad compared with UPRO	Experiment 2: Iliad compared with UCDS
Total number of cases reviewed	100	326
Total number of cases flagged by nurse review	28	88
Number of cases with appropriate care	11	65
Number of cases with quality problems	17	23
Level I (score 1)	10	20
Level II (score 5)	5	3
Level III (score 25)	2	0
DER of reviewed cases	17.0%	7.1%
TPR of nurse review	60.7%	26.1%
Average severity score of cases with problems	5.3	1.5

problem was present, the lowest severity rating should be level I. The category "care appropriate" was meant to be used for cases with no quality problems. In the calculation of DER and TPR for this experiment, a gold standard review marked as "diagnostic error present" and "care appropriate" was counted as a level I problem. A review marked as "diagnostic error uncertain" and "care appropriate" would not have been counted as a quality problem. The large proportion of cases with level I diagnostic errors judged to have had appropriate care suggests that these UPRO physician reviewers felt diagnostic errors to be unimportant if patient management was otherwise acceptable.

To test further our hypothesis that the background of the review physician strongly influenced the outcome of Iliad review, 15 cases in the second experiment were reviewed by the same University of Utah physicians who performed the gold standard Iliad review in the first study and were compared with the gold standard review by the UPRO physicians. These university physicians were blinded to the Iliad gold standard review results by the UPRO physicians. Interrater reliability between the university physicians and the UPRO physicians for Iliad review was calculated.[26,25] Cohen's κ was found to be .44 not statistically significant at $p < .05$ (one-tailed t test). Therefore, there was no agreement between the two groups (Figure 1).

In addition, 17 cases were independently reviewed by a unique UPRO review physician who was not one of the six gold standard reviewers in the experiment. This unique UPRO physician was a pulmonary specialist (but not the university pulmonologist) with teaching responsibilities, whereas the other six UPRO physicians were

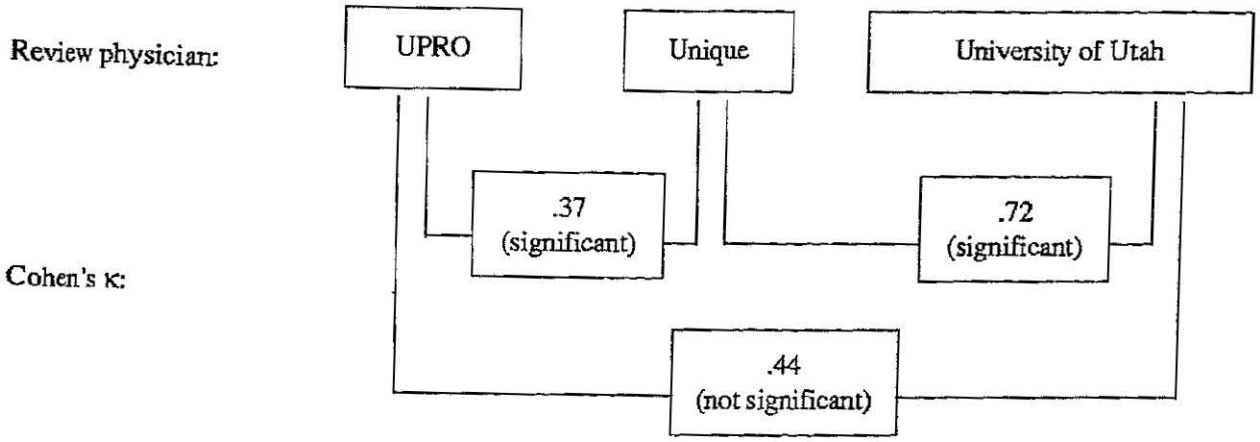

Figure 1. Interrater reliability between different gold standard physician reviewers.

either general internists or family practitioners. Also, this unique physician had been involved in other Iliad research projects. The unique UPRO physician was blinded to the gold standard review results of the other UPRO physicians. Interrater reliability calculated by Cohen's κ was found to be .37, statistically significant ($p < .025$ one-tailed t test).[24,25] This unique UPRO physician also reviewed 7 of the cases reviewed by the university physicians. He was also blinded to the Iliad gold standard review results of the university physicians. Interrater reliability calculated by Cohen's κ was found to be .72, statistically significant ($p < .005$, one-tailed t test). These explorations of interrater reliability are summarized in Figure 1.

Thus there was significant agreement between this unique physician and the UPRO physicians and between this unique physician and the University of Utah physicians but not between the UPRO and University of Utah physicians. The unique physician, sharing background with both groups of physicians, appeared to achieve agreement with each of the two groups, whereas the two groups, which had no common background, were not able to agree with each other. This evidence indirectly supports the conclusion that the type of gold standard reviewers strongly affects the performance of Iliad in diagnostic error detection.

In our experiment, because of financial constraints, we did not study the interrater reliability among the six UPRO review physicians or that among the four University of Utah physicians. It is likely that the physicians within each group agreed with each other because their backgrounds are similar. Our explorations of interrater reliability between groups, however, appears to support the common view that medical opinions vary greatly, as evidenced by the large variation in medical practice documented in the medical literature.[26–29] Although the PRO review was not specifically evaluated regarding interrater reliability among review physicians, PROs were aware that review results varied from PRO to PRO. Physicians also have expressed dissatisfaction with the poor interrater agreement in PRO peer review.[30–33]

Our experiment has shown that a diagnostic expert system such as Iliad can be used effectively as a screening tool to help PRO review nurses flag diagnostic errors, but the end results depend on the review physicians. It is clear from the data presented that diagnostic error detection

in individual chart review by individual physician reviewers may not be reproducible. On the other hand, expert systems such as Iliad can be used to create and tune a diagnostic model based on the consensus opinion of experts from a variety of backgrounds. This diagnostic model can then act as a standard against which diagnoses made by the attending physician can be compared. Chart review by individual physicians can then be substituted by periodic sampling to evaluate and fine-tune the diagnostic model, again by expert consensus opinion. This will drastically reduce the cost and inconsistency of quality review.

REFERENCES

1. Vanagunas, A., et al. "Principles of Quality Assurance." *Quality Review Bulletin* 5 (1979): 4–7.
2. Vanagunas, A. "Quality Assessments: Alternative Approaches." *Quality Review Bulletin* 5 (1979): 8–11.
3. Jessee, W.F. "Quality Assurance Systems. Why Aren't There Any?" *Quality Review Bulletin* 10 (1984): 408–411.
4. Donabedian, A. "Quality Assessment and Assurance: Unity of Purpose, Diversity of Means." *Inquiry* 25 (1988): 173–192.
5. Moore, S., Ryan, C.P. and Murphy, R.A. "PRO Review Part III: Quality Review, Intervention, and Sanction." *Journal of the Medical Association of Georgia* 78 (1989): 683–688.
6. Murphy, R.A., Williams, T.W., and Ryan, C.P. "Peer Review in the 1990s: A Look at the Georgia Medical Care Foundation." *Journal of the Medical Association of Georgia* 78 (1989): 549–551.
7. Utah Peer Review Organization. *Annual Report to Physicians and Health Care Facilities. Review Results April, 1989–March, 1990*. Salt Lake City, Utah: Utah Peer Review Organization, 1990.
8. Mercer, J., and Talbot, I.C. "Clinical Diagnosis: A Post-Mortem Assessment of Accuracy in the 1980s" *Postgraduate Medical Journal* 61 (1985): 713–716.
9. Anderson, E.A., Rolla, B.H., and Key, C.R. "The Sensitivity and Specificity of Clinical Diagnostics during Five Decades." *Journal of the American Medical Association* 261 (1989): 1610–1617.
10. Weinberg, N.S. "The Relation of Medical Problem Solving and Therapeutic Errors to Disease Categories." *Quality Review Bulletin* 15 (1989): 266–272.
11. Williamson, J.W., et al. "Assessing Quality of Ambulatory Care." *Quality Assurance and Utilization Review* 6 (1991): 8–15.
12. Kassirer, J.P., and Kopelman, R.I. "Cognitive Errors in Diagnosis Instantiation, Classification, and Consequences." *American Journal of Medicine* 86 (1989): 433–441.
13. Wilson, J.D., et al. *Harrison's Principles of Internal Medicine*. 12th ed. New York, N.Y.: McGraw Hill, 1991.
14. Adams, I.D., et al. "Computer-Aided Diagnosis of Acute Abdominal Pain: A Multicentre Study." *British Medical Journal* 293 (1986): 800–804.
15. Chassin, R., et al. "Does Inappropriate Use Explain Geographic Variations in the Use of Health Care Services? A Study of Three Procedures." *Journal of the American Medical Association* 258 (1987): 2533–2537.
16. Annual reports from the following peer review organizations: Nevada Peer Review (1990), Puerto Rico Foundation for Medical Care (1990), California Medical Review (1990), New Mexico Medical Review Association (1991), Island Peer Review Organization (New York, 1990), Medical Society of Virginia (1991), Arkansas Foundation for Medical Care (1990), Peer Review Systems (Ohio, 1990), Colorado Foundation for Medical Care (1990), Iowa Foundation for Medical Care (1990), Michigan Peer Review Organization (1990).
17. Peer Review Organization of Washington, *PRO/W Annual Report for Alaska, Idaho and Washington 1991*. Seattle, Wash.: PRO/W, 1991.
18. Stewart, R.D., et al. "A Computer-Assisted Qual-

ity Assurance System for an Emergency Service." *Annals of Emergency Medicine* 14 (1985): 25–29.

19. Lau, L.M., and Warner, H.R. "Performance of a Diagnostic Expert System (Iliad) as a Tool for Quality Assurance." *Computers and Biomedical Research* 25 (1992): 314–323.
20. Warner, H.R., et al. "Iliad as an Expert Consultant To Teach Differential Diagnosis." Presented at the 12th Annual Symposium on Computer Applications in Medical Care, Washington, D.C., November 1988.
21. American Medical Review Research Center (AMRRC). *An Introduction to the Uniform Clinical Data Set*. Washington, D.C.: AMRRC, 1991.
22. Hutton, R.C., Pate, P., and Simmons, G. "The Uniform Clinical Data Set Project." Presented at the 5th Annual Symposium of the American Medical Review Research Center, Boulder, Colo., August 1990.
23. Dawson-Saunders, B., and Trapp, R.G. *Basic and Clinical Biostatistics*. Norwalk, Conn.: Appleton & Lange, 1990.
24. Cohen, J. "A Coefficient of Agreement for Nominal Scales." *Education and Psychological Measurement* 20 (1960): 37–46.
25. Siegel, S., and Castellan, N.J. *Nonparametric Statistics for the Behavioral Sciences*. New York, N.Y.: McGraw-Hill, 1989.
26. Chassin, M.R., et al. "Variations in the Use of Medical and Surgical Services by the Medicare Population." *New England Journal of Medicine* 318(1988): 721–727.
27. Wennberg, J.E., and Cittelsohn, A. "Variations in Medical Care among Small Areas." *Scientific American* 246 (1982): 120–134.
28. Caper, P.S. "Variations in Medical Practice: Implications for Health Policy." *Health Affairs* 3 (1984): 110–129.
29. Wennberg, J.E. "Dealing with Medical Practice Variations: A Proposal for Action." *Health Affairs* 3 (1984): 6–32.
30. Lohr, K., ed. *Medicare: A Strategy for Quality Assurance*. Vol. 1. Washington, D.C.: National Academy Press, 1990.
31. Lohr, K.N., Winkler, J.D., and Brook, R.H. *Peer Review and Technology Assessment in Medicine*. Santa Monica, Calif.: Rand Corporation, 1981.
32. Hershey, N., and Bontempo, L.C. "Assessing Peer Review in the Quest for Improved Medical Services." *Quality Assurance and Utilization Review* 4 (1989): 94–100.
33. Office of Technology Assessment. *The Quality of Medical Care. Information for Consumers*. Washington, D.C.: Government Printing Office, 1988.